

REMARKS

Claims 1-2 and 14-15 stand rejected under 35 U.S.C. 103(a) as unpatentable over Carus '268 in view of McKeown and in further view of Matsubayashi. The references disclosed by examiner do not teach all of the elements of Applicant's claims. And even if one were to assume that the elements are present, a *prima facie* case of obviousness has not been established by the examiner. The examiner has taken references from disparate fields with different objectives and combined aspects of these disparate references motivated solely by the teachings of Applicant. This use of hindsight reconstruction based on art that does not teach the disputed elements fails to make out a proper rejection.

The examiner incorrectly states that McKeown discloses "traversing substrings of the natural-language input in an order determined by the weights assigned to the breakpoints ... to efficiently and accurately identify segment." But, McKeown does not come close to teaching or even suggesting this missing element of claim 1. Assuming for the sake of argument that McKeown can even be said to disclose weighted breakpoints, there clearly is no traversal of any substrings in an order determined by those weights. The examiner points to column 8, lines 24-49 as teaching this claim element. As even the examiner notes, this is a sequential approach to determining segment importance and coverage. Fig. 12 confirms that the approach is sequential. (Fig. 12 is first referenced at column 8, line 66.) The program proceeds, "for each segment..." to score segment significance. The examiner points to nothing in McKeown to suggest proceeding through the scoring algorithm in an order determined by weighting factors. McKeown proceeds sequentially through each segment and makes absolutely no mention of changing the order in response to any weighting factors. Given that neither Carus, McKeown nor Matsubayashi disclose traversing "in an order determined by the weights," claims 1, 2, 14 and 15 should be allowed.

Unlike any of Carus '268, McKeown or Matsubayashi, Applicant sought to solve the problem of segmenting a compound word. This is more of an issue for a language such as German. Consider one of the examples given in Applicant's specification, "Abhaengigkeitsverhaeltnis." This word can be translated as, "[a] state of dependency," or alternatively, "[a] relationship of dependency." In German, "abhaengigkeit," and "verhaeltnis" are words meaning roughly, "dependence," and "relationship," respectively.

However, the linguistically correct way to combine these two words to express a “relationship of dependence,” is to combine them into the compound word given above. In German, the rules of grammar cause extremely long compound words to be formed on a regular basis, many of which may not even have previously existed. It is impossible to predict and store in a lexicon all compound words that can be formed (column 1, paragraph 0004). This is one of the problems that Applicant’s invention addresses. The problem of segmenting compound words is not related to the Asian language lexical analysis and paragraph topical analysis of the cited references.

Carus ‘268 seeks to separate individual words in written Asian language text, a task made more difficult in these languages because the text does not separate them by spaces. Each transition between consecutive characters, according to Carus ‘268, is identified as a link, a breakpoint or an unknown. (col. 5, l. 62-65). Carus ‘268 describes using a sliding window in an orderly manner sliding the window along the text as statistical analysis is conducted. The statistical analysis in Carus ‘268 either finds a breakpoint or does not. There is no disclosure, suggestion or teaching of the assignment of weights to the transitions. All unknown transitions are equally identified as simply unknown. Applicant’s invention which includes the assignment of weights and the use of those weights in traversing substrings is neither taught, disclosed nor suggested by Carus ‘268.

The examiner cites Matsubayashi for its disclosed use of probabilities. Matsubayashi provides head position probabilities and tail position probabilities for characters in a character string, in particular Asian language character strings. These two probabilities for adjacent characters produces a division probability. “When the value of a calculated division probability exceeds a predetermined value (which will be referred to as a division threshold, hereafter), the system performs division of the single character type string thereat.” (Matsubayashi, col. 3, l. 55-59).

Applicant finds no suggestion in Matsubayashi of the use of probabilities for providing an order in which substrings are traversed while seeking to identify linkable components. Rather, Matsubayashi uses the calculations to divide up a character string to extract smaller strings for document searching. The frequency of occurrence of the extracted strings in documents in the databases is used in conducting the search to

identify relevant documents. Applicant finds no suggestion within the references to make use of probabilities as taught by Matsubayashi in the manner claimed and taught by Applicant.

The method of McKeown is fundamentally different from the art of Carus, such that there would not have been a motive to combine the references, and such an attempted combination would make no sense. The fundamental difference between Carus and McKeown, as is explained in detail below, is that Carus seeks to identify words, and McKeown seeks to identify the function of words. Carus deals with correctly identifying individual words in a stream of words containing no visible breaks between words. McKeown takes a larger piece of text that is already broken into words and paragraphs, and then proceeds to analyze the relationship between the paragraphs and the document as a whole, based on the frequency with which certain words appear. This process is similar to that used in a search engine, which determines relevancy of a document to a query based on the frequency with which the query terms appear in the document. This sort of process is completely unrelated to the lexical analysis in Carus. That McKeown is unconcerned with lexical processing is even clearer in light of its specific teachings. McKeown develops its topical map using steps such as merging pronouns with their respective nouns (column 5, lines 11-12), dropping adjectives from noun phrases (column 5, lines 13-19), and filtering out “irrelevant words” (column 5, lines 20-22). These are steps for building a fuzzy, conceptual understanding of a document, not for processing the lexical content of a stream of text. There is no such thing as an “irrelevant” word in lexical parsing. McKeown’s high-level, fuzzy, topical analysis simply bears no relation whatsoever to the problems of either Carus, or of Applicant’s invention.

The examiner states the case for obviousness thus: “Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Carus’ method wherein it is described as above, to provide a method that segments information according to the segment function and the importance, for efficient and accurate segmentation.” The terms “function” and “importance,” taken from McKeown, are concepts from the distant fields of topical analysis (e.g., McKeown) and search engines. By “function,” as used in context in McKeown, is meant the function of a paragraph in a larger text, e.g. a paragraph may function as an introduction, a conclusion,

etc. (column 1, lines 29-34). Again, note that Carus is directed to identifying well-formed words in a stream of text, where the absence of word breaks in the text (a phenomenon common in Asian languages) makes it necessary to do so. Using a method of analyzing previously identified words within paragraphs (such previous identification consisting of nothing more than noting that every string of characters between spaces is a word, and that every string of characters between carriage returns is a paragraph), will not improve a method of identifying words in a stream of text with no spaces. When developing techniques for finding words in a stream of unbroken text, it is irrelevant what techniques will be used to process those words after they are found and extracted.

McKeown, on the other hand, does not address partitioning of a character string at all. Rather, McKeown analyzes a text of sentences and paragraphs and performs topical segmentation on the document and classifies the segments according to function and importance. Applicant respectfully submits there has been no rationale given for one in the art of word break analysis to turn to the art of paragraph segmentation and topical analysis. There is no rationale for combining Carus and McKeown with each other, because they deal with fundamentally different technologies, and because no improvement over Carus could be achieved by adopting the method of McKeown. For these additional reasons, no *prima facie* case of obviousness has been made out. Claims 1, 2, 14 and 15 should be allowed.

Given that none of the cited references disclose traversing “in an order determined by the weights” and that McKeown is totally unrelated to Applicant’s invention or to the Asian language parsing references, Applicant submits that all claims presently in the application are allowable over the art of record and early notice to that effect is respectfully solicited.

Respectfully submitted,

/Robert M. Asher, #30,445/
Robert M. Asher
Reg. No. 30,445
Bromberg & Sunstein LLP
125 Summer Street, 11th Floor
Boston, MA 02110-1618
(617) 443-9292
Attorney for Applicant